Multiscale network-based approaches in bioinformatics and biomedicine

Zheng, H., & Hu, X. (2020). Multiscale network-based approaches in bioinformatics and biomedicine. *Methods*, *173*, 1-2. https://doi.org/10.1016/j.ymeth.2020.01.002

Published in:
Methods

Publication Status:
Published (in print/issue): 15/02/2020

DOI:
10.1016/j.ymeth.2020.01.002

Document Version
Publisher's PDF, also known as Version of record

Methods 173 (2020) 1–2

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Guest Editor's Introduction

Multiscale network-based approaches in bioinformatics and biomedicine

Biological systems are inherently complex in nature across multiple diverse temporal, spatial and functional scales. In recent years, novel network-based approaches have been introduced and developed to deepen the understanding of complex biological systems and human diseases from heterogeneous data types.

Cancer is a complex disease characterized by uncontrolled, uncoordinated, and undesirable growth of abnormal malignant cells. In this special issue, researchers report their progress in different aspects of cancer research using network-based approaches.

Identification of subtypes can improve the understanding of cancer biology and enhances both diagnosis and prognosis, especially provide patients personalized treatment. Mallavarapu et al. [1] proposed a pathway-based deep clustering (PACL) approach that integrates genomics data and biological pathway datasets to investigate the subtypes of complex cancer associated with patients survivals. The framework of the proposed model consists of a gene layer, a pathway layer, two hidden layers, and a cluster layer. The authors claimed two subtypes were discovered in both GBM and ovarian cancer.

Long non-coding RNAs (lncRNAs) play significant roles in modulating gene expression at various levels. Prediction of associations between lncRNAs and disease can help to identify biomarkers for disease diagnosis and treatment. Wang et al. [2] introduce a weighted matrix factorization model on multi-relational data to predict long non-coding RNA-disease associations. A heterogeneous network is built to present inter(intra)-associations between different types of data, including genes, lncRNAs, diseases, drugs and gene ontology terms. Association matrix was ranked, weighted and optimized for the prediction. The proposed multiple heterogeneous networks method was applied to predict the lncRNA-disease association for breast cancer, gastric cancer, and lung cancer. Results demonstrate that the proposed method is capable of selectively integrating multiple inter-relational and intra-relational data sources for a better prediction of the lncRNAs-disease association. Genomic variants are associated with many human diseases including cancers. Spence et al. [3] investigated structural variants between a child and parents and reported a mathematical model and optimization framework for the prediction of structural variants in the context of one parent and one child. The work would be applied to cancer research and beyond.

Yan et al. [4] focused on improving breast cancer histopathological image classification. A new hybrid convolutional and recurrent deep neural network was proposed, which integrates convolutional and recurrent neural networks. The authors concluded that the proposed method outperforms several state-of-the-art algorithms. The authors also released a large dataset of breast cancer pathological images to the scientific community, which will be valuable to other researchers

Interpreting relationships among diseases and symptoms from electronic medical records (EMR) is critical for effective treatments. Guo et al. [5] applied bidirectional Long Short Term Memory networks (BiLSTMs) to model symptom sequence in EMR. Two symptom representation models were combined with BiLSTMS in their study, i.e. term frequency-inverse document frequency (TF-IDF) model for the representation of the relationship between symptoms and diseases; and Word2Vec for the expression of the semantic relationship between symptoms. Results presented in [5] show that BiLSTEMs with TF-IDF and Word2Vec could improve the prediction of disease inference.

Du et al. [6] focused mining knowledge from biomedical question answering data. To address the challenge of small datasets, the authors proposed a hierarchical multi-layer transfer learning model to answer questions from a given contextual passage in the Biomedical field. A gated scaled dot-product attention based neural networks were developed for the Biomedical Question Answering task by the use of transfer learning techniques.

A wide range of algorithms has been researched for the enhancement and detection of curvilinear structures in biomedical image processing. Alharbi et al. [7] introduced a multi-scale concept in the development of a new method, called Multiscale Top-Hat Tensor (MTHT), to enhance the curvilinear. The proposed MTHT was applied to both 2D and 3D images, for example, retina images, and achieved higher specificity and accuracy of the segmentation result when compared to other existing enhancement approaches, although it is sensitive to noise.

Recent years have witnessed the rapid development in human microbiome research. The human microbiome has the potential to affect many aspects of human health and disease. Zhu et al. [8] applied deep random forest to investigate the microbiome-wide associations and proposed an ensemble feature selection method for the identification of microbial biomarkers. The research could help the diagnosis of microbial-related diseases.

Computational tools and frameworks are important for researchers. This special issue also includes a few recently developed or enhanced tools. To fully understand how a glycan moiety will structurally impact the protein surface, BesanÇon et al. [9] developed the Umbrella Visualization method, which takes into account both the protein topology and glycan motions above the protein surface. The visualisation tool provides a graphical method that assigns the glycan intrinsic exibility during a molecular dynamics trajectory and can be used to investigate the protein glycan interactions.

Network topology modeling is a common approach to represent and study the connectivity patterns and functions of a biological system. Subgraph topological patterns, i.e. network motifs, with a similar structure are considered of providing an important function in the biological system. Wang et al. [10] present a new tool to discover network motifs for a given network.

Jiang et al. [11] present a high dimensional multi-omics integration

https://doi.org/10.1016/j.ymeth.2020.01.002

Available online 08 January 2020

tool, IMPRes-Pro, for *in silico* hypothesis generation and testing. The tool integrating proteomics data along with transcriptomics data and constructing a heterogeneous multi-omics network, which can clearly show the relationships among protein-TF-gene. The tool is made available and has been evaluated by the authors with the human metastatic breast cancer dataset.

Network-based approaches have been applied to understand and interpret the complexity of biological systems. We hope this special issue report a set of representative work in this area and provide readers references for further improvement in related areas.

Finally, Guest Editors would like to thank authors for their contributions to this special issue and sharing research findings. We would also like to thank reviewers for providing constructive comments and the journal Editorial members for assisting the editorial work.

References

[1] Mallavarapu, et al., Pathway-based deep clustering for molecular subtyping of cancer, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.06.017.
[2] Wang, et al., Weighted matrix factorization on multi-relational data for LncRNA-disease association prediction, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.06.015.
[3] Spence, et al., Detecting inherited and novel structural variants in low-coverage parent-child sequencing data, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.06.025.
[4] Yan, et al., Breast cancer histopathological image classification using a hybrid deep neural network, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.06.014.
[5] Guo, et al., A disease inference method based on symptom extraction and bidirectional Long Short Term Memory networks, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.07.009.
[6] Du, et al., Deep scaled dot-product attention based domain adaptation model for biomedical question answering, Methods (2020), https://doi.org/10.1016/j.ymeth.2019.06.024.
[7] Alharbi, et al., The multiscale top-hat tensor enables specific enhancement of curvilinear structures in 2D and 3D images, Methods (2020), https://doi.org/10.1016/j.ymeth.2019.05.025.
[8] Zhu, et al., Robust biomarker discovery for microbiome-wide association studies, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.06.012.
[9] Besanon, et al., Umbrella visualization: a method of analysis dedicated to glycan flexibility with UnityMol, Methods (2020), https://doi.org/10.1016/j.ymeth.2019.07.010.
[10] F.S.M. Wang, et al., Fast and scalable network motif discovery for exploring higher-order network organizations, Methods (2020), https://doi.org/10.1016/j.ymeth.2019.07.008.
[11] Jiang, et al., IMPRes-Pro: a high dimensional multiomics integration method for in silico hypothesis generation, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.06.013.

Guest Editors
Huiru (Jane) Zheng*, Xiaohua (Tony) Hu
[a] *School of Computing, Ulster University, UK*
[b] *Drexel University, USA*
E-mail addresses: h.zheng@ulster.ac.uk (H.J. Zheng), xh29@drexel.edu (X.T. Hu).

* Corresponding author.